GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 7, 2003, 16:13:48 ; Search time 14 Seconds
(without alignments)
956.918 Million cell updates/sec

Title: US-09-988-292-16
Perfect score: 1759
Sequence: 1 MAYVPAPGYQPTYNPTLPYY.....FQRVDTLEIQGDVTLSYVQI 323

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1759	100.0	323	1	LEG4_HUMAN	P56470 homo sapien
2	1419	80.7	323	1	LEG4_PIG	Q29058 sus scrofa
3	1367.5	77.7	324	1	LEG4_RAT	P38552 rattus norv
4	1227	69.8	301	1	LEG6_MOUSE	O54891 mus musculu
5	590	33.5	353	1	LEG9_MOUSE	O08573 mus musculu
6	588	33.4	354	1	LEG9_RAT	P97840 rattus norv
7	574	32.6	355	1	LEG9_HUMAN	O00182 homo sapien
8	531.5	30.2	316	1	LEG8_HUMAN	O00214 homo sapien
9	513.5	29.2	316	1	LEG8_MOUSE	Q9jl15 mus musculu
10	507.5	28.9	316	1	LEG8_RAT	Q62665 rattus norv
11	445.5	25.3	283	1	LEG1_HAECO	O44126 haemonchus
12	422	24.0	279	1	LE32_CAEEL	P36573 caenorhabdi
13	400	22.7	297	1	LE33_CAEEL	Q09581 caenorhabdi
14	338.5	19.2	336	1	LEGC_HUMAN	Q96dt0 homo sapien
15	283.5	16.1	135	1	LEG7_MOUSE	O54974 mus musculu
16	277.5	15.8	135	1	LEG7_RAT	P97590 rattus norv
17	275.5	15.7	144	1	LEG5_RAT	P47967 rattus norv
18	272.5	15.5	135	1	LEG7_HUMAN	P47929 homo sapien
19	272.5	15.5	244	1	LEG3_CRILO	P47953 cricetulus
20	264.5	15.0	261	1	LEG3_RAT	P08699 rattus norv
21	263.5	15.0	263	1	LEG3_MOUSE	P16110 mus musculu
22	255	14.5	295	1	LEG3_CANFA	P38486 canis famil
23	241.5	13.7	249	1	LEG3_HUMAN	P17931 homo sapien
24	240	13.6	241	1	LEG3_RABIT	P47845 oryctolagus
25	174.5	9.9	134	1	LEG6_CHICK	P23668 gallus gall
26	172.5	9.8	134	1	LEG4_CHICK	P07583 gallus gall
27	171.5	9.7	141	1	LPPL_HUMAN	Q05315 homo sapien
28	162	9.2	129	1	LEG_ELEEL	P08520 electrophor
29	157	8.9	180	1	LEC8_CAEEL	Q09610 caenorhabdi
30	156	8.9	134	1	LEG1_HUMAN	P09382 homo sapien
31	153.5	8.7	130	1	LEG2_MOUSE	Q9cqw5 mus musculu
32	153	8.7	134	1	LEG1_SHEEP	P81184 ovis aries
33	153	8.7	1216	1	YPX5_CAEEL	Q09277 caenorhabdi
34	152	8.6	179	1	LEC7_CAEEL	Q09605 caenorhabdi
35	151	8.6	134	1	LEG1_MOUSE	P16045 mus musculu
36	150	8.5	134	1	LEG1_BOVIN	P11116 bos taurus
37	147	8.4	134	1	LEG1_CRIGR	P48538 cricetulus
38	146	8.3	134	1	LEG1_RAT	P11762 rattus norv
39	140.5	8.0	130	1	LEG2_RAT	Q9z144 rattus norv
40	133.5	7.6	123	1	LEG2_PIG	Q29373 sus scrofa
41	132.5	7.5	132	1	LEG2_HUMAN	P05162 homo sapien
42	127.5	7.2	134	1	LEG1_BUFAR	P56217 bufo arenar
43	116	6.6	135	1	LEG1_CONMY	P26788 conger myri
44	102.5	5.8	135	1	LEG2_CONMY	Q9yic2 conger myri
45	97.5	5.5	105	1	LEGX_MOUSE	P11946 mus musculu

ALIGNMENTS

RESULT 1
LEG4_HUMAN
ID LEG4_HUMAN STANDARD; PRT; 323 AA.
AC P56470;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE (L36LBP) (Antigen NY-CO-27).
GN LGALS4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric carcinoma;
RA Kato S.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RX MEDLINE=97454307; PubMed=9310382;
RA Rechreche H., Mallo G.V., Montalto G., Dagorn J.C., Iovanna J.L.;
RT "Cloning and expression of the mRNA of human galectin-4, an S-type
RT lectin down-regulated in colorectal cancer.";
RL Eur. J. Biochem. 248:225-230(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97306339; PubMed=9162064;
RA Huflejt M.E., Jordan E.T., Gitt M.A., Barondes S.H., Leffler H.;
RT "Strikingly different localization of galectin-3 and galectin-4 in
RT human colon adenocarcinoma T84 cells. Galectin-4 is localized at sites
RT of cell adhesion.";
RL J. Biol. Chem. 272:14294-14303(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Skin;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC ---

CC ---
DR EMBL; AB006781; BAA22165.1; -.

```
DR   EMBL; AF014838; AAC51763.1; -.
DR   EMBL; U82953; AAB86590.1; -.
DR   EMBL; BC003661; AAH03661.1; -.
DR   EMBL; BC005146; AAH05146.1; -.
DR   HSSP; P47929; 1BKZ.
DR   Genew; HGNC:6565; LGALS4.
DR   MIM; 602518; -.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    152       GALAPTIN 1.
FT   DOMAIN      153    177       LINKER.
FT   DOMAIN      178    323       GALAPTIN 2.
FT   BINDING     256    262       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   323 AA;  35941 MW;  E79EC0A9AB3990EF CRC64;

  Query Match             100.0%;  Score 1759;  DB 1;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-137;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323


RESULT 2
LEG4_PIG
ID   LEG4_PIG       STANDARD;      PRT;   323 AA.
AC   Q29058; Q29296;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE   (L36LBP).
GN   LGALS4.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Tongue;
RX   MEDLINE=95081129; PubMed=7989350;
RA   Chiu M.L., Parry D.A.D., Feldman S.R., Klapper D.G., O'Keefe E.J.;
RT   "An adherens junction protein is a member of the family of lactose-
RT   binding lectins.";
RL   J. Biol. Chem. 269:31770-31776(1994).
RN   [2]
RP   SEQUENCE OF 1-140 FROM N.A.
RC   TISSUE=Small intestine;
RX   MEDLINE=96327607; PubMed=8672129;
RA   Winteroe A.K., Fredholm M., Davies W.;
RT   "Evaluation and characterization of a porcine small intestine cDNA
RT   library: analysis of 839 clones.";
RL   Mamm. Genome 7:509-517(1996).
CC   -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC       SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79303; CAA55884.1; -.
DR   EMBL; F14653; CAA23179.1; -.
DR   HSSP; P47929; 1BKZ.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; FALSE_NEG.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    152       GALAPTIN 1.
FT   DOMAIN      153    177       LINKER.
FT   DOMAIN      178    323       GALAPTIN 2.
FT   BINDING     256    262       BETA-GALACTOSIDE (BY SIMILARITY).
FT   CONFLICT     98     99       AP -> GA (IN REF. 2).
FT   CONFLICT    126    126       H -> T (IN REF. 2).
SQ   SEQUENCE   323 AA;  35852 MW;  728C761712B29343 CRC64;

  Query Match             80.7%;  Score 1419;  DB 1;  Length 323;
  Best Local Similarity   80.5%;  Pred. No. 2.1e-109;
  Matches  260;  Conservative   25;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||:|||||||||||||||:|||||| ||||||||||:||||||||||||| ||:||
Db       1 MAFVPAPGYQPTYNPTLPYYKPIPGGLRVGMSVYIQGVANEHMKRFFVNFVVGQGPGADV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||: | ||||:||:||||||:|  ||||| :|| |||||||||:|
Db      61 AFHFNPRFDGWDKVVFNSQQDGKWGNEEKKRSMPFRKAPAFELVIMVLPEHYKVVVNGDP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           |||:|||:|:|:||||||||| ||||||||||    ||   | |||||  :||  :||
Db     121 FYEFGHRIPVQLVTHLQVDGDLTLQSINFIGGQPAPSPGPMPNPGYPGPGKHNQQPCNLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||| |||||||||  |||||| |||||:|||||:|||  ||||||||||:|||||||
Db     181 CMEGAPTFNPPVPYKTRLQGGLVARRTIVIKGYVPPSGKSLVINFKVGSSGDVALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           :  | ||||| ||| ||:||:| : ||| |||:||||||||||||||||||||:||
Db     241 LTEGIVVRNSYLNGKWGAEERKSSFNPFAPGQYFDLSIRCGLDRFKVYANGQHLFDFSHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           || || ||||||||||||||||||
Db     301 LSNFQGVDTLEIQGDVTLSYVQI 323


RESULT 3
LEG4_RAT
ID   LEG4_RAT       STANDARD;      PRT;   324 AA.
AC   P38552;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Galectin-4 (Lactose-binding lectin 4) (L-36 lactose binding protein)
DE   (L36LBP).
```

```
GN   LGALS4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=93194902; PubMed=8449956;
RA   Oda Y., Herrmann J., Gitt M., Turck C.W., Burlingame A.L.,
RA   Barondes S.H., Leffler H.;
RT   "Soluble lactose-binding lectin from rat intestine with two different
RT   carbohydrate-binding domains in the same peptide chain.";
RL   J. Biol. Chem. 268:5929-5939(1993).
CC   -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC       SUGARS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FULL-LENGTH FORM IN SMALL
CC       AND LARGE INTESTINE AND STOMACH BUT WAS NOT DETECTED IN OTHER
CC       TISSUES INCLUDING LUNG, LIVER, KIDNEY, AND SPLEEN.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73553; AAA41505.1; -.
DR   PIR; A46631; A46631.
DR   HSSP; P47929; 1BKZ.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    152       GALAPTIN 1.
FT   DOMAIN      153    177       LINKER.
FT   DOMAIN      178    324       GALAPTIN 2.
FT   BINDING     257    263       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   324 AA;  36347 MW;  478024D7322AFE7B CRC64;

  Query Match             77.7%;  Score 1367.5;  DB 1;  Length 324;
  Best Local Similarity   76.6%;  Pred. No. 3.6e-105;
  Matches  249;  Conservative   33;  Mismatches   40;  Indels    3;  Gaps    2;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           |||||||||||||||||| :||||||:||||:||||:| ::|:|| ||| |||| |:|:
Db       1 MAYVPAPGYQPTYNPTLPYKRPIPGGLSVGMSIYIQGIAKDNMRRFHVNFAVGQDEGADI 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           |||||||||||||||||:| |:|| ||:|:||||:||  |||||:|::|||||||| |
Db      61 AFHFNPRFDGWDKVVFNTMQSGQWGKEEKKKSMPFQKGHHFELVFMVMSEHYKVVVNGTP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMM--PPYPGPGHCHQQLNS 178
           |||||||||||||||||||:||||||:||||   | |  |  | ||  |:   |:||
Db     121 FYEYGHRLPLQMVTHLQVDGDLELQSINFLGGQPAASQYPGTMTIPAYPSAGYNPPQMNS 180

Qy     179 LPTMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHIN 238
           || | ||| ||||||| | ||||||||||||||||| || |:  |||||||:|||| |:|
Db     181 LPVMAGPPIFNPPVPYVGTLQGGLTARRTIIIKGYVLPTAKNLIINFKVGSTGDIAFHMN 240

Qy     239 PRMGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFA 298
           ||:|:  ||||| :||||||||:|| :|||| |||||||||| |||||:|||||||||:
Db     241 PRIGD-CVVRNSYMNGSWGSEERKIPYNPFGAGQFFDLSIRCGTDRFKVFANGQHLFDFS 299

Qy     299 HRLSAFQRVDTLEIQGDVTLSYVQI 323
           ||  |||||| |||:||:|||||||
Db     300 HRFQAFQRVDMLEIKGDITLSYVQI 324
```

```
RESULT 4
LEG6_MOUSE
ID   LEG6_MOUSE     STANDARD;      PRT;   301 AA.
AC   O54891; O88352;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Galectin-6.
GN   LGALS6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=98112847; PubMed=9446608;
RA   Gitt M.A., Colnot C., Poirier F., Nani K.J., Barondes S.H.,
RA   Leffler H.;
RT   "Galectin-4 and galectin-6 are two closely related lectins expressed
RT   in mouse gastrointestinal tract.";
RL   J. Biol. Chem. 273:2954-2960(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=98112848; PubMed=9446609;
RA   Gitt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Barondes S.H.,
RA   Leffler H.;
RT   "Sequence, structure, and chromosomal mapping of the mouse Lgals6
RT   gene, encoding galectin-6.";
RL   J. Biol. Chem. 273:2961-2970(1998).
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF026799; AAC04508.1; -.
DR   EMBL; AF026796; AAC04508.1; JOINED.
DR   EMBL; AF026797; AAC04508.1; JOINED.
DR   EMBL; AF026798; AAC04508.1; JOINED.
DR   EMBL; AF026794; AAC27244.1; -.
DR   HSSP; P17931; 1A3K.
DR   MGD; MGI:107535; Lgals6.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    151       GALAPTIN 1.
FT   DOMAIN      152    160       LINKER.
FT   DOMAIN      160    301       GALAPTIN 2.
FT   CONFLICT    154    154       A -> V (IN REF. 2; AAC27244).
SQ   SEQUENCE   301 AA;  34112 MW;  9A4DD09944EDFAB9 CRC64;

  Query Match             69.8%;  Score 1227;  DB 1;  Length 301;
  Best Local Similarity   70.3%;  Pred. No. 1.1e-93;
  Matches  227;  Conservative   32;  Mismatches   42;  Indels   22;  Gaps    3;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           |||||||||||||||||| :||||||:|||| |||| | |:|:|| ||| |||| |:||
Db       1 MAYVPAPGYQPTYNPTLPYKRPIPGGLSVGMSFYIQGTAKENMRRFHVNFAVGQDDGADV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           |||||||||||||||||| | |:|| || | ||||:||  |||||:|: ||||||||:|
```

```
Db      61 AFHFNPRFDGWDKVVFNTKQSGRWGKEEEK-SMPFQKGKHFELVFMVMPEHYKVVVNGSP 119

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           |||||||||:||||||||||||:|||||| | ||                         |
Db     120 FYEYGHRLPVQMVTHLQVDGDLELQSINFFGVQPAE--------------------TKYP 159

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           | ||| ||| :|| | |||| | |||||||||| || |:|||||:|||| |||||||||
Db     160 AMTGPPVFNPCLPYVGALQGGFTVRRTIIIKGYVLPTAKTFAINFRVGSSEDIALHINPR 219

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           :|:  :|||| :|||||:||: : :|||||||||||||||:|||||:||| |||:|:||
Db     220 IGD-CLVRNSYMNGSWGTEERMVAYNPFGPGQFFDLSIRCGMDRFKVFANGIHLFNFSHR 278

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
             | ::::|||| ||:||||| |
Db     279 FQALRKINTLEINGDLTLSYVHI 301


RESULT 5
LEG9_MOUSE
ID   LEG9_MOUSE     STANDARD;      PRT;   353 AA.
AC   O08573; O08572;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Galectin-9.
GN   LGALS9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Small intestine, and Kidney;
RX   MEDLINE=97190351; PubMed=9038233;
RA   Wada J., Kanwar Y.S.;
RT   "Identification and characterization of galectin-9, a novel beta-
RT   galactoside-binding mammalian lectin.";
RL   J. Biol. Chem. 272:6078-6086(1997).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=97298141; PubMed=9153289;
RA   Wada J., Ota K., Kumar A., Wallner E.I., Kanwar Y.S.;
RT   "Developmental regulation, expression, and apoptotic potential of
RT   galectin-9, a beta-galactoside binding lectin.";
RL   J. Clin. Invest. 99:2452-2461(1997).
CC   -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
CC       EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-
CC       CLASSICAL SECRETORY PATHWAY.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC       HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ACCENTUATED EXPRESSION IN LIVER AND THYMUS OF
CC       EMBRYO, DETECTED IN EMBRYONIC HEART, BRAIN, LUNG, LIVER, AND
CC       KIDNEY. HIGHLY EXPRESSED IN ADULT THYMUS, SMALL INTESTINE, AND
CC       LIVER, AND TO A LESSER EXTENT IN LUNG, KIDNEY, SPLEEN, CARDIAC,
CC       AND SKELETAL MUSCLE. BARELY DETECTABLE IN BRAIN AND RETICULOCYTE.
CC       THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
CC   -!- DEVELOPMENTAL STAGE: THE EXPRESSION INCREASED WITH SUCCESSIVE
CC       STAGES OF EMBRYONIC DEVELOPMENT.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U55061; AAB51190.1; -.
DR   EMBL; U55060; AAB51189.1; -.
DR   HSSP; P17931; 1A3K.
DR   MGD; MGI:109496; Lgals9.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat; Alternative splicing.
FT   DOMAIN        1    147       GALAPTIN 1.
FT   DOMAIN      148    204       LINKER.
FT   DOMAIN      205    353       GALAPTIN 2.
FT   BINDING      81     87       BETA-GALACTOSIDE (BY SIMILARITY).
FT   BINDING     285    291       BETA-GALACTOSIDE (BY SIMILARITY).
FT   VARSPLIC    148    178       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   353 AA;  40036 MW;  B54036F6E280C531 CRC64;

  Query Match             33.5%;  Score 590;  DB 1;  Length 353;
  Best Local Similarity   38.7%;  Pred. No. 2.7e-41;
  Matches  135;  Conservative   46;  Mismatches  126;  Indels   42;  Gaps    9;

Qy      11 PTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDVAFHFNPRFDG 70
             |  || :|:  || |||  |: | :||      :|| |||      |:|:||||||:
Db       9 PYINPIIPFTGPIQGGLQEGLQVTLQGTTKSFAQRFVVNF-QNSFNGNDIAFHFNPRFEE 67

Qy      71 WDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPL 130
           || || | |:|| |||| |||:|| ||| |:|    :||:||   | :| ||:|
Db      68 GGYVVCNTKQNGQWGPEERKMQMPFQKGMPFELCFLVQRSEFKVMVNKKFFVQYQHRVPY 127

Qy     131 QMVTHLQVDGDLQLQSINFIG-------------GQPLR----PQGPPMMPPYPGPGH- 171
           :|  : | | |:|  | |                 ||::    |:|         | |
Db     128 HLVDTIAVSGCLKLSFITFQNSAAPVQHVFSTLQFSQPVQFPRTPKGRKQKTQNFRPAHQ 187

Qy     172 ----------CHQ---QLNSLPTMEGPPTFNP----PVPYFGRLQGGLTARRTIIIKGYV 214
                      |    |: | | :  ||   |     :|::  :  ||   ::|:| | |
Db     188 APMAQTTIHMVHSTPGQMFSTPGI--PPVVYPTPAYTIPFYTPIPNGLYPSKSIMISGNV 245

Qy     215 PPTGKSFAINFKVGSSGDIALHINPRMGNGTVVRNSLLNGSWGSEEKKIT-HNPFGPGQF 273
           |   | || : |  |||| |:|||     ||||: :| ||| ||: :   ||  ||
Db     246 LPDATRFHINLRCG--GDIAFHLNPRFNENAVVRNTQINNSWGQEERSLLGRMPFSRGQS 303

Qy     274 FDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRVDTLEIQGDVTLSYVQ 322
           | : | |    ||| ||||: :: |||   | ::|||: ||: |::||
Db     304 FSVWIICEGHCFKVAVNGQHMCEYYHRLKNLQDINTLEVAGDIQLTHVQ 352


RESULT 6
LEG9_RAT
ID   LEG9_RAT       STANDARD;      PRT;   354 AA.
AC   P97840; O08588; O35866;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Galectin-9 (36 kDa beta-galactoside binding lectin) (Urate
DE   transporter/channel) (UAT).
GN   LGALS9.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG AND SHORT FORMS).
RC   STRAIN=Sprague-Dawley; TISSUE=Kidney, and Small intestine;
RX   MEDLINE=97190351; PubMed=9038233;
RA   Wada J., Kanwar Y.S.;
RT   "Identification and characterization of galectin-9, a novel beta-
RT   galactoside-binding mammalian lectin.";
RL   J. Biol. Chem. 272:6078-6086(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX   MEDLINE=97150769; PubMed=8995305;
RA   Leal-Pinto E., Tao W., Rappaport J., Richardson M., Knorr B.A.,
```

```
RA   Abramson R.G.;
RT   "Molecular cloning and functional reconstitution of a urate
RT   transporter/channel.";
RL   J. Biol. Chem. 272:617-625(1997).
CC   -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
CC       EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY
CC       SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS
CC       CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE
CC       METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC
CC       TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND
CC       GASTROINTESTINAL URATE EXCRETION. HIGHLY SELECTIVE TO THE ANION
CC       URATE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-
CC       CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC       HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE
CC       SMALL INTESTINE.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U59462; AAB51192.1; -.
DR   EMBL; U72741; AAB68592.1; -.
DR   EMBL; U67958; AAB48591.1; -.
DR   HSSP; P17931; 1A3K.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat; Alternative splicing; Ion transport.
FT   DOMAIN        1    147       GALAPTIN 1.
FT   DOMAIN      148    205       LINKER.
FT   DOMAIN      206    354       GALAPTIN 2.
FT   BINDING      81     87       BETA-GALACTOSIDE (BY SIMILARITY).
FT   BINDING     286    292       BETA-GALACTOSIDE (BY SIMILARITY).
FT   VARSPLIC    148    179       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   354 AA;  39946 MW;  6574F960B2EAF37C CRC64;

  Query Match             33.4%;  Score 588;  DB 1;  Length 354;
  Best Local Similarity   38.8%;  Pred. No. 4e-41;
  Matches  138;  Conservative   43;  Mismatches  123;  Indels   52;  Gaps   10;

Qy      10 QPTY-NPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDVAFHFNPRF 68
           || | || :|:   | ||| |: : :||       |  |||  |   |:|:||||||||
Db       7 QPPYMNPVIPFTGIIQGGLQNGLQITLQGTVHPFPNRIAVNFQTGFS-GNDIAFHFNPRF 65

Qy      69 DGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRL 128
           :   || || | |||| ||||  |||:||  ||| |:|    :||:|| | | :| ||:
Db      66 EEGGYVVCNTKQNGKWGPEERKMQMPFQKGMPFELCFLVQRSEFKVMVNKNFFVQYSHRV 125

Qy     129 PLQMVTHLQVDGDLQLQSINFIG-------------------------GQPLRPQG--P 160
           |  :|  : | | | |  |||                          |:  | ||  |
Db     126 PYHLVDTISVSGCLHLSFINFQNSTAAPVQPVFSTMQFSQPVQFPRMPKGRKQRTQGFQP 185

Qy     161 PMMPP---------YPGPGHCHQQLNSLPTMEGPPTFNP----PVPYFGRLQGGLTARRT 207
           :  |         :  ||    |: | | :  ||   |     :|:|  :  |    ::
Db     186 ALQAPVAQTIIHTVHSIPG----QMLSTPGI--PPMAYPTPAYTIPFFTSIPNGFYPSKS 239

Qy     208 IIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGNGTVVRNSLLNGSWGSEEKKIT-HN 266
           | | | | |  | | || : |  |||| |:|||     ||||: :| ||| ||: :
Db     240 INISGVVLPDAKRFHINLRCG--GDIAFHLNPRFNEKVVVRNTQINNSWGPEERSLPGRM 297

Qy     267 PFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRVDTLEIQGDVTLSYVQ 322
           ||  || | : | |    |||  :|||: :: |||     ::|||: ||: |::||

Db     298 PFNRGQSFSVWILCEGHCFKVAVDGQHICEYYHRLKNLPDINTLEVAGDIQLTHVQ 353


RESULT 7
LEG9_HUMAN
ID   LEG9_HUMAN     STANDARD;      PRT;   355 AA.
AC   O00182; O14532; O75028; Q9NQ58;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Galectin-9 (HOM-HD-21) (Ecalectin).
GN   LGALS9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RC   TISSUE=Spleen;
RX   MEDLINE=97197815; PubMed=9045665;
RA   Tuereci O., Schmitt H., Fadle N., Pfreundschuh M., Sahin U.;
RT   "Molecular definition of a novel human galectin which is immunogenic
RT   in patients with Hodgkin's disease.";
RL   J. Biol. Chem. 272:6416-6422(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RC   TISSUE=Gastric carcinoma;
RA   Kato S.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RX   MEDLINE=98307937; PubMed=9642261;
RA   Matsumoto R., Matsumoto H., Seki M., Hata M., Asano Y., Kanegasaki S.,
RA   Stevens R.L., Hirashima M.;
RT   "Human ecalectin, a variant of human galectin-9, is a novel eosinophil
RT   chemoattractant produced by T lymphocytes.";
RL   J. Biol. Chem. 273:16976-16984(1998).
RN   [4]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RA   Akiyama S.;
RT   "Homo sapiens galectin-9 (LGALS9) / ecalectin gene, exon 2 through
RT   11.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RA   Graessler J., Spitzenberger F., Schroeder H.E.;
RT   "Genomic organization of the human galectin-9 gene.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
CC       EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE
CC       SHORT ISOFORM ACTS AS AN EOSINOPHIL CHEMOATTRACTANT.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC       HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC
CC       TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49107; CAA88922.1; -.
DR   EMBL; AB006782; BAA22166.1; -.
DR   EMBL; AB005894; BAA31542.1; -.
DR   EMBL; AB040130; BAB83625.1; -.
DR   EMBL; AB040129; BAB83625.1; JOINED.
DR   EMBL; AB040130; BAB83624.1; -.
```

```
DR   EMBL; AB040129; BAB83624.1; JOINED.
DR   EMBL; AJ288083; CAB93851.1; -.
DR   EMBL; AJ288084; CAB93851.1; JOINED.
DR   EMBL; AJ288085; CAB93851.1; JOINED.
DR   EMBL; AJ288086; CAB93851.1; JOINED.
DR   EMBL; AJ288087; CAB93851.1; JOINED.
DR   EMBL; AJ288088; CAB93851.1; JOINED.
DR   EMBL; AJ288089; CAB93851.1; JOINED.
DR   EMBL; AJ288090; CAB93851.1; JOINED.
DR   HSSP; P17931; 1A3K.
DR   Genew; HGNC:6570; LGALS9.
DR   MIM; 601879; -.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat; Alternative splicing.
FT   DOMAIN        1    148       GALAPTIN 1.
FT   DOMAIN      149    206       LINKER.
FT   DOMAIN      207    355       GALAPTIN 2.
FT   BINDING      82     88       BETA-GALACTOSIDE (BY SIMILARITY).
FT   BINDING     287    293       BETA-GALACTOSIDE (BY SIMILARITY).
FT   VARSPLIC    149    180       MISSING (IN SHORT ISOFORM).
FT   CONFLICT      5      5       G -> S (IN REF. 3).
FT   CONFLICT     48     48       N -> D (IN REF. 5).
FT   CONFLICT     79     81       NGS -> KGR (IN REF. 5).
FT   CONFLICT     88     88       K -> R (IN REF. 1).
FT   CONFLICT     89     89       T -> M (IN REF. 5).
FT   CONFLICT    135    135       S -> F (IN REF. 1).
FT   CONFLICT    270    270       P -> L (IN REF. 1).
FT   CONFLICT    313    313       E -> G (IN REF. 1).
FT   CONFLICT    326    326       L -> V (IN REF. 5).
FT   CONFLICT    341    341       R -> K (IN REF. 5).
SQ   SEQUENCE   355 AA;  39518 MW;  4748C22FCAFA536A CRC64;

  Query Match             32.6%;  Score 574;  DB 1;  Length 355;
  Best Local Similarity   38.2%;  Pred. No. 5.7e-40;
  Matches  134;  Conservative   52;  Mismatches  121;  Indels   44;  Gaps   10;

Qy      11 PTYNPTLPYYQPIPGGLNVGMSVYIQG-VASEHMKRFFVNFVVGQDPGSDVAFHFNPRFD 69
           |  :| :|:   | |||  |: : : | | |    || |||  |   |:|:|||||||:
Db       9 PYLSPAVPFSGTIQGGLQDGLQITVNGTVLSSSGTRFAVNFQTGFS-GNDIAFHFNPRFE 67

Qy      70 GWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLP 129
             || || | | || ||||  |||:||  |:| |:| :  :||:|||  | :| ||:|
Db      68 DGGYVVCNTRQNGSWGPEERKTHMPFQKGMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP 127

Qy     130 LQMVTHLQVDGDLQLQSINFIGGQ-------------------PLRPQG-----PPMMPP 165
             |  : |:| :||  |:|   :                     | ||:|     | : |
Db     128 FHRVDTISVNGSVQLSYISFQNPRTVPVQPAFSTVPFSQPVCFPPRPRGRRQKPPGVWPA 187

Qy     166 YPGP-----GHCHQ----QLNSLPTMEGPPTFNP----PVPYFGRLQGGLTARRTIIIKG 212
           | |       |  |    |: | | :  ||    |    |:|:   : |||   ::|:: |
Db     188 NPAPITQTVIHTVQSAPGQMFSTPAI--PPMMYPHPAYPMPFITTILGGLYPSKSILLSG 245

Qy     213 YVPPTGKSFAINFKVGSSGDIALHINPRMGNGTVVRNSLLNGSWGSEEKKITHN-PFGPG 271
           | |: : | ||  : |   || |:|||     ||||: :: ||||||: :    ||  |
Db     246 TVLPSAQRFHIN--LCSGNHIAFHLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRG 303

Qy     272 QFFDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRVDTLEIQGDVTLSYVQ 322
           | | : | |     ||  :|||||:: |||     :: ||: ||: |::||
Db     304 QSFSVWILCEAHCLKVAVDGQHLFEYYHRLRNLPTINRLEVGGDIQLTHVQ 354


RESULT 8
LEG8_HUMAN
ID   LEG8_HUMAN     STANDARD;      PRT;   316 AA.
AC   O00214; O15215; Q9UP34; Q9UEZ6; Q9UP33; Q9UP32; Q9H584; Q9H585;
AC   Q96B92;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Galectin-8 (Gal-8) (Prostate carcinoma tumor antigen 1) (PCTA-1)
DE   (Po66 carbohydrate-binding protein) (Po66-CBP).
GN   LGALS8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=96293510; PubMed=8692978;
RA   Su Z.-Z., Lin J., Shen R., Fisher P.E., Goldstein N.I., Fisher P.B.;
RT   "Surface-epitope masking and expression cloning identifies the human
RT   prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin
RT   gene family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:7252-7257(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hippocampus;
RA   Hadari Y.R., Eisenstein M., Zakut R., Zick Y.;
RT   "Galectin-8: on the road from structure to function.";
RL   Trends Glycosci. Glycotechnol. 9:103-112(1997).
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Lung carcinoma;
RA   Brichory F., Bidon N., Desrues B., Bourguet P., Le Pennec J.P.,
RA   Dazord L.;
RT   "Molecular cloning of a beta-galactoside-binding lectin related to
RT   galectin-8 and identified in human lung carcinoma.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maier C., Haeussler J., Roesch K., Moschgath E., Vogel W.;
RT   "Genomic organization and expression of the human galectin-8 gene.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438187; PubMed=10980616;
RA   Gopalkrishnan R.V., Roberts T., Tuli S., Kang D., Christiansen K.A.,
RA   Fisher P.B.;
RT   "Molecular characterization of prostate carcinoma tumor antigen-1,
RT   PCTA-1, a human galectin-8 related gene.";
RL   Oncogene 19:4405-4416(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colorectal carcinoma;
RA   Lahm H., Siebert H.-C., Andre S., Hoeflich A., Diehl D., Sordat B.,
RA   Kaltner H., Wolf E., Gabius H.-J.;
RT   "Coca (Colorectal carcinoma-derived) galectin-8 variant I full-length
RT   cDNA from a human colorectal carcinoma cell line.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Cobley V.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Brain, and Skin;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/I (shown here), 2 and 3; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: UNIQUITOUS. SELECTIVE EXPRESSION BY PROSTATE
CC       CARCINOMAS VERSUS NORMAL PROSTATE AND BENIGN PROSTATIC
CC       HYPERTROPHY.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L78132; AAB51605.1; -.
DR   EMBL; X91790; CAA62904.1; ALT_INIT.
DR   EMBL; AF074000; AAD45402.1; -.
DR   EMBL; AF074001; AAD45403.1; -.
DR   EMBL; AF074002; AAD45404.1; -.
DR   EMBL; AF193806; AAF19370.1; ALT_INIT.
DR   EMBL; AF193805; AAF19370.1; JOINED.
DR   EMBL; AF342815; AAK16735.1; ALT_INIT.
DR   EMBL; AL136105; CAC15946.1; ALT_INIT.
DR   EMBL; AL136105; CAC15947.1; ALT_INIT.
DR   EMBL; BC015818; AAH15818.1; -.
DR   EMBL; BC016486; AAH16486.1; ALT_INIT.
DR   HSSP; P17931; 1A3K.
DR   Genew; HGNC:6569; LGALS8.
DR   MIM; 606099; -.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 1.
KW   Galaptin; Lectin; Repeat; Antigen; Alternative splicing; Polymorphism.
FT   DOMAIN        1    153       GALAPTIN 1.
FT   DOMAIN      154    184       LINKER.
FT   DOMAIN      185    316       GALAPTIN 2.
FT   BINDING     248    254       BETA-GALACTOSIDE (BY SIMILARITY).
FT   VARSPLIC    182    182       L -> LPSNRGGDISKIAPRTVYTKSKDSTVNHTLTCTKIP
FT                                PMNYVSK (IN ISOFORM 2).
FT   VARSPLIC    182    182       L -> LQTVSPSWDLQGHGSETFCSVLWTRVFLEIAFCRPI
FT                                GLTVASFQ (IN ISOFORM 3).
FT   VARIANT      18     18       Y -> F.
FT                                /FTId=VAR_012990.
FT   VARIANT      35     35       C -> R.
FT                                /FTId=VAR_009710.
FT   VARIANT      55     55       V -> M.
FT                                /FTId=VAR_012991.
FT   CONFLICT     97     99       KRE -> QKEK (IN REF. 2).
FT   CONFLICT    111    111       D -> A (IN REF. 2).
FT   CONFLICT    170    170       S -> V (IN REF. 1).
FT   CONFLICT    183    183       S -> R (IN REF. 2, 6, 7 AND 8).
FT   CONFLICT    203    203       K -> Q (IN REF. 1).
SQ   SEQUENCE   316 AA;  35538 MW;  1724C36085A0510A CRC64;

  Query Match             30.2%;  Score 531.5;  DB 1;  Length 316;
  Best Local Similarity   37.6%;  Pred. No. 1.5e-36;
  Matches  120;  Conservative   42;  Mismatches  132;  Indels   25;  Gaps    4;

Qy      13 YNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQD--PGSDVAFHFNPRFDG 70
           ||| :||   ||  |: |  : | |        || |:   |    | :||||||||||
Db      12 YNPVIPYVGTIPDQLDPGTLIVICGHVPSDADRFQVDLQNGSSVKPRADVAFHFNPRFKR 71

Qy      71 WDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPL 130
             :| |||   ||| ||     |||:  :||:| :|| : ::| |||     ||||:
Db      72 AGCIVCNTLINEKWGREEITYDTPFKREKSFEIVIMVLKDKFQVAVNGKHTLLYGHRIGP 131

Qy     131 QMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP----TMEGPP 186
           : :  | : | : : || |     |:                  | :||    : |  |
Db     132 EKIDTLGIYGKVNIHSIGFSFSSDLQ---------------STQASSLELTEISRENVP 175

Qy     187 TFNPP---VPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGN 243
               |   :|:  ||    :   ||:::|| |    ||| ::   | | |||||:|||:
Db     176 KSGTPQLSLPFAARLNTPMGPGRTVVVKGEVNANAKSFNVDLLAGKSKDIALHLNPRLNI 235

Qy     244 GTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHRLSA 303
             |||| |  ||| ||: ||  || || :|:: | | :  |||  || |  :: ||
Db     236 KAFVRNSFLQESWGEEERNITSFPFSPGMYFEMIIYCDVREFKVAVNGVHSLEYKHRFKE 295

Qy     304 FQRVDTLEIQGDVTLSYVQ 322
             :||||| ||: |  |:
Db     296 LSSIDTLEINGDIHLLEVR 314
```

```
RESULT 9
LEG8_MOUSE
ID   LEG8_MOUSE     STANDARD;      PRT;   316 AA.
AC   Q9JLI5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Galectin-8 (LGALS-8).
GN   LGALS8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RA   Maier C., Haeussler J., Roesch K., Moschgath E., Haeusler J.,
RA   Vogel W.;
RT   "The human Lgals-8 gene: genomic sequence and expression of the
RT   prostate carcinoma tumour antigen (PCTA-1) and the Po66 carbohydrate
RT   binding protein.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POSSESSES SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AF218069; AAF27645.1; -.
DR   HSSP; P17931; 1A3K.
DR   MGD; MGI:1928481; Lgals8.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 1.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    153       GALAPTIN 1.
FT   DOMAIN      154    184       LINKER.
FT   DOMAIN      185    316       GALAPTIN 2.
FT   BINDING     248    254       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   316 AA;  36161 MW;  11A20309AEF52C69 CRC64;

  Query Match             29.2%;  Score 513.5;  DB 1;  Length 316;
  Best Local Similarity   36.0%;  Pred. No. 4.6e-35;
  Matches  111;  Conservative   49;  Mismatches  137;  Indels   11;  Gaps    3;

Qy      13 YNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQD--PGSDVAFHFNPRFDG 70
           ||| :||   |   |  | : |:|   :  :|| |:| :|    | :||||||||||
Db      12 YNPIIPYVGTITEQLKPGSLIVIRGHVPKDSERFQVDFQLGNSLKPRADVAFHFNPRFKR 71

Qy      71 WDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPL 130
             :| |||   ||| ||     |||:|  :||:||:||   ::| |||     | ||:
Db      72 SSCIVCNTLTQEKWGWEEITYDMPFRKEKSFEIVFMVLKNKFQVAVNGRHVLLYAHRISP 131

Qy     131 QMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLPTMEGPPTFNP 190
           : :  : : | : : || |     |:      :           |:|    :: |
Db     132 EQIDTVGIYGKVNIHSIGFRFSSDLQSMETSALG--------LTQINR-ENIQKPGKLQL 182

Qy     191 PVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGNGTVVRNS 250
           :|:  ||   :   ||::||| |    :|| ::   | : |||||:|||:     ||||
Db     183 SLPFEARLNASMGPGRTVVIKGEVNTNARSFNVDLVAGKTRDIALHLNPRLNVKAFVRNS 242

Qy     251 LLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRVDTL 310
             |  :|| ||: ||  ||  | :|:: | | :  |||  || |  :: ||      :|||
Db     243 FLQDAWGEEERNITCFPFSSGMYFEMIIYCDVREFKVAINGVHSLEYKHRFKDLSSIDTL 302

Qy     311 EIQGDVTL 318
```

```
          : ||: |
Db    303 SVDGDIRL 310


RESULT 10
LEG8_RAT
ID   LEG8_RAT       STANDARD;      PRT;   316 AA.
AC   Q62665;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Galectin-8 (30 kDa S-type lectin) (RL-30).
GN   LGALS8.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=95155445; PubMed=7852431;
RA   Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y.;
RT   "Galectin-8. A new rat lectin, related to galectin-4.";
RL   J. Biol. Chem. 270:3447-3453(1995).
CC   -!- FUNCTION: POSSESSES SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, CARDIAC MUSCLE,
CC       LUNG, AND BRAIN.
CC   -!- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN WHOLE EMBRYOS, HIGH LEVELS
CC       IN ADULT TISSUES.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09824; AAA66359.1; -.
DR   HSSP; P17931; 1A3K.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 1.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    153       GALAPTIN 1.
FT   DOMAIN      154    184       LINKER.
FT   DOMAIN      185    316       GALAPTIN 2.
FT   BINDING     248    254       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   316 AA;  36038 MW;  C04B766CFE913D59 CRC64;

  Query Match             28.9%;  Score 507.5;  DB 1;  Length 316;
  Best Local Similarity   34.7%;  Pred. No. 1.4e-34;
  Matches  111;  Conservative   45;  Mismatches  129;  Indels   35;  Gaps    3;

Qy      13 YNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQD--PGSDVAFHFNPRFDG 70
           ||||:||   |   |  |  : |:|   :  :|| |:|  |    | :||||||||||
Db      12 YNPTIPYVSTITEQLKPGSLIVIRGHVPKDSERFQVDFQHGNSLKPRADVAFHFNPRFKR 71

Qy      71 WDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPL 130
           : :| |||   ||| ||   |||:|  :||:| :||   : | |||     | ||:
Db      72 SNCIVCNTLTNEKWGWEEITHDMPFRKEKSFEIVIMVLKNKFHVAVNGKHILLYAHRINP 131

Qy     131 QMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLPT--------- 181
           : :  | : | : : || |                          | |: |
Db     132 EKIDTLGIFGKVNIHSIGF--------------------RFSSDLQSMETSTLGLTQIS 170

Qy     182 ---MEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHIN 238
              ::    :  :|:  ||   :   ||:::|| |     || ::   | | |||||:|

Db     171 KENIQKSGKLHLSLPFEARLNASMGPGRTVVVKGEVNTNATSFNVDLVAGRSRDIALHLN 230

Qy     239 PRMGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFA 298
           ||:     |||| |  :|| ||: ||  ||  | :|:: | | :  |||  || |  ::
Db     231 PRLNVKAFVRNSFLQDAWGEEERNITCFPFSSGMYFEMIIYCDVREFKVAVNGVHSLEYK 290

Qy     299 HRLSAFQRVDTLEIQGDVTL 318
           ||      :||| : ||: |
Db     291 HRFKDLSSIDTLAVDGDIRL 310


RESULT 11
LEG1_HAECO
ID   LEG1_HAECO     STANDARD;      PRT;   283 AA.
AC   O44126;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   32 kDa beta-galactoside-binding lectin (Galectin 1).
GN   GAL-1.
OS   Haemonchus contortus (Barber pole worm).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC   Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX   NCBI_TaxID=6289;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Moredun; TISSUE=Gut;
RX   MEDLINE=20067137; PubMed=10599080;
RA   Newlands G.F.J., Skuce P.J., Knox D.P., Smith S.K., Smith W.D.;
RT   "Cloning and characterization of a beta-galactoside-binding protein
RT   (galectin) from the gut of the gastrointestinal nematode parasite
RT   Haemonchus contortus.";
RL   Parasitology 119:483-490(1999).
CC   -!- FUNCTION: BINDS GALACTOSE.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF036098; AAB88823.1; -.
DR   HSSP; P17931; 1A3K.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat.
FT   DOMAIN        1    150       GALAPTIN 1.
FT   DOMAIN      151    283       GALAPTIN 2.
FT   BINDING     217    223       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   283 AA;  32530 MW;  84D319BF6F383750 CRC64;

  Query Match             25.3%;  Score 445.5;  DB 1;  Length 283;
  Best Local Similarity   33.2%;  Pred. No. 1.5e-29;
  Matches  104;  Conservative   52;  Mismatches  114;  Indels   43;  Gaps    8;

Qy      13 YNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNF--VVGQDPGSDVAFHFNPRFDG 70
           ||  :||   :   :  | :: ::|   :  :|| :|         |:||  | : |||
Db      11 YNKPVPYRSLLQEKIEPGQTLIVKGSTIDESQRFTINLHSKSADFSGNDVPLHISVRFDE 70

Qy      71 WDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPL 130
           |||  ||   |:|| |||| |:| ||| :|::      : :::|::   | :| |||||
Db      71 -GKVVMNTFANGEWGKEERK-SLPIKKGDSFDIRIRAHDDRFQIVIDQKEFKDYEHRLPL 128

Qy     131 QMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLPTMEGPPTFNP 190
             :||| :|||| |  ::: ||:
Db     129 TSITHLSIDGDLYLNHVHW-GGKYY---------------------------------- 152

Qy     191 PVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGNGTVVRNS 250
```

```
          ||||   :  |   :|::| | |   | | ||  :  :|||||| |||     |:||:
Db    153 PVPYESGIASGFPIDKTLLIFGTVEKKAKRFNINL-LRRNGDIALHFNPRFDEKAVIRNA 211

Qy    251 LLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRVDTL 310
          |   ||:||::   ||  |  |||:|:    |::: ||:   ||||     :  |
Db    212 LAANEWGNEERE-GKMPFEKGVGFDLAIKNEAYAFQIFVNGERFTSFAHRQDP-NDISGL 269

Qy    311 EIQGDVTLSYVQI 323
          :||||: |: :||
Db    270 QIQGDIELTGIQI 282


RESULT 12
LE32_CAEEL
ID   LE32_CAEEL     STANDARD;      PRT;   279 AA.
AC   P36573;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   32 kDa beta-galactoside-binding lectin (32 kDa GBP).
GN   W09H1.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92348399; PubMed=1639789;
RA   Hirabayashi J., Satoh M., Kasai K.-I.;
RT   "Evidence that Caenorhabditis elegans 32-kDa beta-galactoside-binding
RT   protein is homologous to vertebrate beta-galactoside-binding lectins.
RT   cDNA cloning and deduced amino acid sequence.";
RL   J. Biol. Chem. 267:15485-15490(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97476274; PubMed=9334250;
RA   Arata Y., Hirabayashi J., Kasai K.-I.;
RT   "Structure of the 32-kDa galectin gene of the nematode Caenorhabditis
RT   elegans.";
RL   J. Biol. Chem. 272:26669-26677(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Smye R.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 202-220.
RX   MEDLINE=92348337; PubMed=1639749;
RA   Hirabayashi J., Satoh M., Ohyama Y., Kasai K.-I.;
RT   "Purification and characterization of beta-galactoside-binding
RT   proteins from Caenorhabditis elegans.";
RL   J. Biochem. 111:553-555(1992).
CC   -!- FUNCTION: BINDS GALACTOSE.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94671; AAB87718.1; -.
DR   EMBL; AB000802; BAA22942.1; -.
DR   EMBL; Z82081; CAB04959.1; -.
DR   PIR; S27798; S27798.
DR   PIR; PX0062; PX0062.
DR   PIR; A42846; A42846.
DR   HSSP; P17931; 1A3K.
DR   WormPep; W09H1.6a; CE16576.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 2.
KW   Galaptin; Lectin; Repeat; Multigene family.
FT   DOMAIN       1    146       GALAPTIN 1.
FT   DOMAIN     147    279       GALAPTIN 2.
FT   BINDING    213    219       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   279 AA;  31809 MW;  ED9AE22AE37571DA CRC64;

  Query Match             24.0%;  Score 422;  DB 1;  Length 279;
  Best Local Similarity   32.0%;  Pred. No. 1.3e-27;
  Matches  101;  Conservative   53;  Mismatches  118;  Indels   44;  Gaps    9;

Qy     10 QPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNF--VVGQDPGSDVAFHFNPR 67
          :|   | :||   :       | :: ::|   :  :|| :|         |:||  | : |
Db      5 EPKSYP-VPYRSVLQEKFEPGQTLIVKGSTIDESQRFTINLHSKTADFSGNDVPLHVSVR 63

Qy     68 FDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHR 127
          ||   |:| |:   |:|| |||| | | ||| :|::      : ::::|:   | :| ||
Db     64 FDE-GKIVLNSFSNGEWGKEERK-SNPIKKGDSFDIRIRAHDDRFQIIVDHKEFKDYEHR 121

Qy    128 LPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLPTMEGPPT 187
          ||| ::|| :|||| |  ::: ||:
Db    122 LPLSSISHLSIDGDLYLNHVHW-GGKYY-------------------------------- 148

Qy    188 FNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGNGTVV 247
             ||||   |  ||   ::::: | |    | | :|  :  :|||: | |||     |:
Db    149 ---PVPYESGLANGLPVGKSLLVFGTVEKKAKRFHVNL-LRKNGDISFHFNPRFDEKHVI 204

Qy    248 RNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHRLSAFQRV 307
          ||||    ||:||::   |||  |  ||| |:     |:|: ||:    ||||      :
Db    205 RNSLAANEWGNEERE-GKNPFEKGVGFDLVIQNEEYAFQVFVNGERYISFAHRADPHD-I 262

Qy    308 DTLEIQGDVTLSYVQI 323
          |:| ||: || :||
Db    263 AGLQISGDIELSGIQI 278


RESULT 13
LE33_CAEEL
ID   LE33_CAEEL     STANDARD;      PRT;   297 AA.
AC   Q09581; Q9GNP6;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   32 kDa beta-galactoside-binding lectin lec-3 (32 kDa GBP).
GN   LEC-3 OR ZK892.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hirabayashi J., Hayama K., Kasai K.;
RT   "Novel galectins found in C. elegans.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Lloyd C.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Jones S.J.M.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS GALACTOSE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB038504; BAB11969.1; -.
DR   EMBL; Z48638; CAA88570.2; -.
DR   HSSP; P47929; 1BKZ.
DR   WormPep; ZK892.1; CE24743.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 2.
DR   SMART; SM00276; GLECT; 2.
DR   PROSITE; PS00309; GALAPTIN; 1.
KW   Galaptin; Lectin; Repeat; Multigene family.
FT   DOMAIN        1    137       GALAPTIN 1.
FT   DOMAIN      138    297       GALAPTIN 2.
FT   BINDING     224    230       BETA-GALACTOSIDE (BY SIMILARITY).
SQ   SEQUENCE   297 AA;  33641 MW;  A48FCED3D8644082 CRC64;

  Query Match             22.7%;  Score 400;  DB 1;  Length 297;
  Best Local Similarity   31.5%;  Pred. No. 8.8e-26;
  Matches  102;  Conservative   52;  Mismatches  110;  Indels   60;  Gaps   11;

Qy      17 LPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQD----PGSDVAFHFNPRFDGWD 72
           :||   :   :  | :: |:|   :  ||| :|  : :|     |:||  | : |||
Db       9 IPYRSKLTERIEPGQTLIIRGKTIDESKRFNIN--LHKDSPDFSGNDVPLHLSIRFDE-G 65

Qy      73 KVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPLQM 132
           |:|:|    | || ||| :: | |||  |::        ::| :|      : ||:||
Db      66 KIVYNAYTKGTWGKEERAKN-PIKKGDDFDIRIRAHDSKFQVSINHKEVKNFEHRIPLNS 124

Qy     133 VTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLPTMEGPPTFNPPV 192
           |:|| :|||: |  : : ||:                                    ||
Db     125 VSHLSIDGDVVLNHVQW-GGKYY----------------------------------PV 148

Qy     193 PY-FGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPRMGN-------- 243
           ||  |    ||   :|::: |      | | ||  :  :|||||| |||
Db     149 PYESGIAADGLVPGKTLVVYGTPEKKAKKFNINL-LKKNGDIALHFNPRFDEKANGFMCA 207

Qy     244 ----GTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAH 299
               |:||||||:|| ||:||::   |||     ||| ||     |::: ||:    :||
Db     208 KPTPGSVVRNSLVNGEWGNEERE-GKNPFERLTAFDLEIRNEEFAFQIFVNGERFASYAH 266

Qy     300 RLSAFQRVDTLEIQGDVTLSYVQI 323
           |:     :  |:||||: |: :|:
Db     267 RVDPHD-IAGLQIQGDIELTGIQV 289


RESULT 14
LEGC_HUMAN
ID   LEGC_HUMAN     STANDARD;      PRT;   336 AA.
AC   Q96DT0; Q96DS9; Q9NZ02; Q96PR9; Q9H259; Q9H258;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Galectin-12 (Galectin-related inhibitor of proliferation).
GN   LGALS12 OR GRIP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D).
RC   TISSUE=Adipose tissue;
RX   MEDLINE=21423969; PubMed=11435439;
RA   Hotta K., Funahashi T., Matsukawa Y., Takahashi M., Nishizawa H.,
RA   Kishida K., Matsuda M., Kuriyama H., Kihara S., Nakamura T.,
RA   Tochino Y., Bodkin N.L., Hansen B.C., Matsuzawa Y.;
RT   "Galectin-12, an adipose-expressed galectin-like molecule possessing
RT   apoptosis-inducing activity.";
RL   J. Biol. Chem. 276:34089-34097(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS B; E AND F).
RC   TISSUE=Retina;
RX   MEDLINE=21283005; PubMed=11283015;
RA   Yang R.-Y., Hsu D.K., Yu L., Ni J., Liu F.-T.;
RT   "Cell cycle regulation by galectin-12, a new member of the galectin
RT   superfamily.";
RL   J. Biol. Chem. 276:20252-20260(2001).
CC   -!- FUNCTION: Binds lactose.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 6 isoforms; a (shown here), b/GRIP1a
CC       c, d, e/1, and f/2; are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Not widely expressed. Predominantly expressed
CC       in adipose tissue.
CC   -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC       DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF244974; AAK77328.1; -.
DR   EMBL; AF244975; AAK77329.1; -.
DR   EMBL; AF244976; AAK77330.1; -.
DR   EMBL; AF244977; AAK77331.1; -.
DR   EMBL; AF310686; AAG40863.1; -.
DR   EMBL; AF310687; AAG40864.1; -.
DR   EMBL; AF222695; AAF34677.1; -.
DR   Genew; HGNC:15788; LGALS12.
DR   MIM; 606096; -.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 1.
DR   SMART; SM00276; GLECT; 1.
DR   PROSITE; PS00309; GALAPTIN; FALSE_NEG.
KW   Galaptin; Lectin; Repeat; Nuclear protein; Alternative splicing.
FT   DOMAIN       47    183       GALAPTIN 1.
FT   DOMAIN      194    336       GALAPTIN 2.
FT   VARSPLIC      1     22       MISSING (IN ISOFORM E AND ISOFORM F).
FT   VARSPLIC      1     61       MISSING (IN ISOFORM B AND ISOFORM D).
FT   VARSPLIC     74     74       H -> HS (IN ISOFORM F).
FT   VARSPLIC    200    208       MISSING (IN ISOFORM C AND ISOFORM D).
SQ   SEQUENCE   336 AA;  37541 MW;  F35D0A0CE509E795 CRC64;

  Query Match             19.2%;  Score 338.5;  DB 1;  Length 336;
  Best Local Similarity   30.7%;  Pred. No. 1.1e-20;
  Matches   98;  Conservative   46;  Mismatches  136;  Indels   39;  Gaps    8;

Qy      11 PTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQD--PGSDVAFHFNPRF 68
           | ::| :||   | |||: |  | :|||       || |:|  |    |  |:|||||||
Db      41 PVFHPVVPYVTTIFGGLHAGKMVMLQGVVPLDAHRFQVDFQCGCSLCPRPDIAFHFNPRF 100

Qy      69 DGW-DKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHR 127
              |: ||| ||:|  | |   :  ::|::| ::|:   |  || |||  |  : :|
Db     101 HTTKPHVICNTLHGGRWQREARWPHLALRRGSSFLILFLFGNEEVKVSVNGQHFLHFRYR 160

Qy     128 LPLQMVTHLQVDGDLQLQSINFIGGQPL------RPQGPPMMPPYPGPGHCHQQLNSLPT 181
           |||  |  | : ||: :::: |:   |        | | |                |  |
Db     161 LPLSHVDTLGIFGDILVEAVGFLNINPFVEGSREYPAGHPF------------LLMSPR 207

Qy     182 MEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPRM 241
           :|       ||    | ||:  : ||::| |    | | :: :    | | |    :
Db     208 LE--------VPCSHALPQGLSPGQVIIVRGLVLQEPKHFTVSLR-----DQAAHAPVTL 254

Qy     242 GNGTVVRNSLLNGSWGSEEKKITHNP--FGPGQFFDLSIRCGLDRFKVYANGQHLFDFAH 299
                 ||  :||:   |  | | :||:: :       |:  ||| |    :
Db     255 RASFADRTLAWISRWG--QKKLISAPFLFYPQRFFEVLLLFQEGGLKLALNGQGLGATSM 312

Qy     300 RLSAFQRVDTLEIQGDVTL 318
             | :::  | | | | |
Db     313 NQQALEQLRELRISGSVQL 331
```

```
RESULT 15
LEG7_MOUSE
ID   LEG7_MOUSE     STANDARD;      PRT;   135 AA.
AC   O54974;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Galectin-7.
GN   LGALS7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=98362649; PubMed=9697310;
RA   Magnaldo T., Fowlis D., Darmon M.;
RT   "Galectin-7, a marker of all types of stratified epithelia.";
RL   Differentiation 63:159-168(1998).
CC   -!- FUNCTION: COULD BE INVOLVED IN CELL-CELL AND/OR CELL-MATRIX
CC       INTERACTIONS NECESSARY FOR NORMAL GROWTH CONTROL (BY SIMILARITY).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MAY BE SECRETED BY A NON-CLASSICAL SECRETORY
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF038562; AAB92566.1; -.
DR   HSSP; P47929; 2GAL.
DR   MGD; MGI:1316742; Lgals7.
DR   InterPro; IPR001079; Galectin.
DR   Pfam; PF00337; Gal-bind_lectin; 1.
DR   SMART; SM00276; GLECT; 1.
DR   PROSITE; PS00309; GALAPTIN; 1.
KW   Galaptin; Lectin.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   BINDING      69     75       BETA-GALACTOSIDE (POTENTIAL).
SQ   SEQUENCE   135 AA;  15042 MW;  9C8E3CBBF36C5013 CRC64;

  Query Match             16.1%;  Score 283.5;  DB 1;  Length 135;
  Best Local Similarity   43.7%;  Pred. No. 1.2e-16;
  Matches   55;  Conservative   24;  Mismatches   46;  Indels    1;  Gaps    1;

Qy      23 IPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDVAFHFNPRFDGWDKVVFNTLQGG 82
           :| |: ||  : |:|:  :   || || : |:: |:| | ||||| |   :||||| | |
Db       9 LPQGVRVGTVMRIRGLVPDQAGRFHVNLLCGEEQGADAALHFNPRLD-TSEVVFNTKQQG 67

Qy      83 KWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLPLQMVTHLQVDGDL 142
           ||| |||   :||::|  ||:: |   | :| ||  : :  : ||||   |  ::| ||:
Db      68 KWGREERGTGIPFQRGQPFEVLLIATEEGFKAVVGDDEYLHFHHRLPPARVRLVEVGGDV 127

Qy     143 QLQSIN 148
           || |:|
Db     128 QLHSLN 133


Search completed: January  7, 2003, 16:18:07
Job time : 15 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 7, 2003, 16:13:23 ; Search time 38 Seconds
(without alignments)
1132.630 Million cell updates/sec

Title: US-09-988-292-16
Perfect score: 1759
Sequence: 1 MAYVPAPGYQPTYNPTLPYY.....FQRVDTLEIQGDVTLSYVQI 323

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1759	100.0	323	18	AAW11841	Human galectin-4-l
2	1759	100.0	323	18	AAW06551	Human colon specif
3	1759	100.0	323	19	AAW56503	Human galectin 8.
4	1759	100.0	323	19	AAW46883	Protein sequence e
5	1759	100.0	323	22	AAM24512	C892P predicted am
6	1759	100.0	323	23	AAO15235	Human galectin-4 (
7	1759	100.0	323	23	AAU76211	Human colon specif
8	1759	100.0	329	23	AAO15241	Human galectin-4 (
9	1755	99.8	351	22	AAG73728	Human colon cancer
10	1475.5	83.9	280	19	AAW61367	Human galectin-4.
11	1458	82.9	268	22	AAG75625	Human colon cancer
12	576	32.7	323	21	AAY56802	Human eosinophil c
13	576	32.7	323	23	ABB77853	Amino acid sequenc
14	576	32.7	323	23	ABP61494	Human NF-kB activa
15	575.5	32.7	341	22	AAB97171	Rainbow trout gale
16	574	32.6	355	20	AAW85664	Galectin-9 like pr
17	574	32.6	355	20	AAY06997	Galectin-9 protein
18	574	32.6	355	23	ABB77852	Amino acid sequenc
19	574	32.6	378	22	AAE13847	Human lung tumour-
20	569	32.3	311	19	AAW56504	Human galectin 9.
21	569	32.3	311	23	ABB77854	Amino acid sequenc
22	569	32.3	329	23	AAU97036	Human bladder canc
23	534.5	30.4	358	21	AAY87404	Human PCTA-1 splic
24	534.5	30.4	358	22	AAB85031	Protein encoded by
25	531.5	30.2	316	21	AAY87403	Human PCTA-1 splic
26	531.5	30.2	316	22	AAB85030	Protein encoded by
27	530.5	30.2	317	19	AAW56505	Human galectin 10.
28	530.5	30.2	368	21	AAY87405	Human PCTA-1 splic
29	530.5	30.2	368	22	AAB85032	Protein encoded by
30	527.5	30.0	317	17	AAW03519	Prostate carcinoma
31	513.5	29.2	316	21	AAY87406	Murine PCTA-1, SEQ
32	507.5	28.9	316	16	AAR75702	Rat galectin-8. R
33	454	25.8	300	23	AAU97824	Human cell membran
34	410.5	23.3	264	20	AAW88106	Human galectin-8 p
35	410.5	23.3	264	22	AAG67687	Amino acid sequenc
36	400.5	22.8	94	23	ABP04847	Human ORFX protein
37	383	21.8	278	16	AAR72597	Ostertagia protect
38	352	20.0	503	22	ABB65681	Drosophila melanog
39	338.5	19.2	314	23	AAM48755	Human adipose tiss
40	338.5	19.2	336	21	AAY70029	Human extracellula
41	327	18.6	168	21	AAB11899	Human colon tumour
42	327	18.6	168	22	AAM24496	Colon tumour relat
43	312	17.7	609	20	AAW82726	Adenovirus pACTSsi
44	312	17.7	609	20	AAW82727	Adenovirus Pig4KN
45	312	17.7	750	20	AAW82728	Adenovirus PNSPig4

ALIGNMENTS

RESULT 1
AAW11841
ID AAW11841 standard; Protein; 323 AA.
XX
AC AAW11841;
XX
DT 20-OCT-1997 (first entry)
XX
DE Human galectin-4-like protein.
XX
KW galectin-4; lactose binding; stomach; intestine; reagent; drug;
KW research; carbohydrate.
XX
OS Homo sapiens.
XX
PN WO9703190-A1.
XX
PD 30-JAN-1997.
XX
PF 09-JUL-1996; 96WO-JP01899.
XX
PR 11-JUL-1995; 95JP-0174778.
XX
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
PI Kamata K, Kato S, Sekine S, Yamaguchi T;
XX
DR WPI; 1997-119046/11.
DR N-PSDB; AAT59539.
XX
PT Recombinant human galectin 4-like protein and gene - has
PT lactose-binding ability, used in research involving carbohydrate(s)
XX

```
PS   Claim 1; Page 13-16; 25pp; Japanese.
XX
CC   This sequence is a human galectin 4 (a lactose-binding protein)-like
CC   protein. The protein, which is expressed specifically in the human
CC   stomach and intestines, binds lactose and is useful as a drug and as a
CC   reagent for research involving carbohydrates.
XX
SQ   Sequence   323 AA;

  Query Match             100.0%;  Score 1759;  DB 18;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-158;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323


RESULT 2
AAW06551
ID   AAW06551 standard; Protein; 323 AA.
XX
AC   AAW06551;
XX
DT   13-MAR-1997  (first entry)
XX
DE   Human colon specific gene CSG9 complete polypeptide.
XX
KW   Colon specific gene; CSG9; colon cancer; metastasis; diagnosis;
KW   therapy; antibody; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO9639419-A1.
XX
PD   12-DEC-1996.
XX
PF   06-JUN-1995;   95WO-US07289.
XX
PR   06-JUN-1995;   95WO-US07289.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Yu G;
XX
DR   WPI; 1997-043054/04.
DR   N-PSDB; AAT45888.
XX
PT   Human colon specific genes and their expression products - detection
PT   of which, in non-colon tissue samples, can be used as indication of
PT   colon cancer metastasis
XX
PS   Claim 8; Fig 9; 60pp; English.
XX
CC   Novel polypeptides (AAW06545-53) are encoded by cDNA clones (see also
CC   AAT45880-92) corresponding to 13 human colon specific genes,
CC   designated CSG1, CSG2, etc., that are primarily expressed in
CC   tissues derived from the colon.  Recombinant CSG polypeptides can
CC   be produced in transformed host cells.  They are useful diagnostic
CC   markers for colon cancer and for colon cancer metastasis and can
CC   also be used to screen for (ant)agonist cpds. of therapeutic or
CC   diagnostic value.  Antibodies raised against the colon-specific
CC   polypeptides may be used to target colon cancer cells or as part
CC   of a colon cancer vaccine.
XX
SQ   Sequence   323 AA;

  Query Match             100.0%;  Score 1759;  DB 18;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-158;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323


RESULT 3
AAW56503
ID   AAW56503 standard; Protein; 323 AA.
XX
AC   AAW56503;
XX
DT   14-SEP-1998  (first entry)
XX
DE   Human galectin 8.
XX
KW   Galectin 8; lectin; human; autoimmune disease;
KW   inflammatory disease; asthma; allergy; melanoma; renal astrocytoma;
KW   Hodgkin disease; breast cancer; ovary cancer; prostate cancer;
KW   bone cancer; liver cancer; lung cancer; pancreas cancer;
KW   spleen cancer; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          55..101
FT                   /note= "antigenic region"
FT   Region          137..162
FT                   /note= "antigenic region"
FT   Region          180..193
FT                   /note= "antigenic region"
FT   Region          216..266
FT                   /note= "antigenic region"
XX
PN   WO9815624-A1.
XX
```

```
PD   16-APR-1998.
XX
PF   09-OCT-1997;   97WO-US18261.
XX
PR   09-OCT-1996;   96WO-US16565.
PR   09-OCT-1996;   96US-0028093.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Gentz RL,  Ni J,  Ruben SM;
XX
DR   WPI; 1998-240812/21.
DR   N-PSDB; AAV29785.
XX
PT   Galectin 8, 9, 10 and 10SV polynucleotides - used for treating
PT   cancer, autoimmune diseases, inflammatory diseases, asthma, and
PT   allergic diseases
XX
PS   Claim 9; Fig 1; 118pp; English.
XX
CC   This polypeptide comprises novel human galectin 8.  Its amino acid
CC   sequence was deduced from a human adult small intestine cDNA clone
CC   (see AAV29785).  Galectin 8 appears to be mainly expressed in cells
CC   of the colon and small intestine.  Novel galectins 8, 9, 10 and 10
CC   splice variant (10SV) are claimed (see AAW56503-06).  These can be
CC   obtained using a claimed method by expression in recombinant host
CC   cells.  The galectins are used in a claimed method for treating a
CC   cell growth disorder, especially cancer, autoimmune diseases,
CC   inflammatory diseases, asthma, and allergic diseases, in a mammal.
CC   They are especially useful for diagnosis of melanoma, renal
CC   astrocytoma, Hodgkin disease, and breast, ovarian, prostate, bone,
CC   liver, lung, pancreatic and splenic cancers.  The invention further
CC   relates to screening methods for identifying agonists and
CC   antagonists of galectin 8, 9, 20 or 10SV activity, and to
CC   diagnostic methods involving estimating levels of galectin 8, 9, 10
CC   or 10SV protein or mRNA.
XX
SQ   Sequence   323 AA;

  Query Match             100.0%;  Score 1759;  DB 19;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-158;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323


RESULT 4
AAW46883
ID   AAW46883 standard; Protein; 323 AA.
XX
AC   AAW46883;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Protein sequence encoded by a colon-specific gene.
XX
KW   Colon-specific gene; probe; detection; expression; human;
KW   diagnostic assay; colon cancer; antibody; screening.
XX
OS   Homo sapiens.
XX
PN   US5733748-A.
XX
PD   31-MAR-1998.
XX
PF   06-JUN-1995;   95US-0469667.
XX
PR   06-JUN-1995;   95US-0469667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen C,  Yu G;
XX
DR   WPI; 1998-229823/20.
DR   N-PSDB; AAV16676.
XX
PT   Colon-specific nucleic acids - useful as probes for detecting colon
PT   cancer micrometastases
XX
PS   Claim 1; Fig 9; 51pp; English.
XX
CC   AAW46876-80 and W4682-85 represent proteins encoded by colon-specific
CC   genes. The polynucleotides encoding these proteins can be used
CC   as probes to detect expression of the corresponding human genes,
CC   e.g. in diagnostic assays for detecting micrometastases of colon cancer.
CC   Recombinant cells containing the polynucleotides can be used to
CC   produce the proteins, in order that antibodies can be raised and
CC   used in further screening or diagnostics.
XX
SQ   Sequence   323 AA;

  Query Match             100.0%;  Score 1759;  DB 19;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-158;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323


RESULT 5
AAM24512
ID   AAM24512 standard; Protein; 323 AA.
XX
```

AC AAM24512;
XX
DT 12-OCT-2001 (first entry)
XX
DE C892P predicted amino acid sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 436-437; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 323 AA;

Query Match 100.0%; Score 1759; DB 22; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.6e-158;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
          |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323
```

RESULT 6
AAO15235
ID AAO15235 standard; Protein; 323 AA.
XX
AC AAO15235;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human galectin-4 (Cln114) protein.
XX
KW Human; gastrointestinal cancer; stomach cancer; small intestine cancer;
KW colon cancer; gastrointestinal specific gene; GSG; galectin-4; Cln114;
KW carbonic anhydrase I; Cln115; gastrointestinal cancer marker.
XX
OS Homo sapiens.
XX
PN US2002042088-A1.
XX
PD 11-APR-2002.
XX
PF 09-MAR-2001; 2001US-0802674.
XX
PR 09-MAR-2000; 2000US-188061P.
XX
PA (MACI/) MACINA R A.
PA (PIDE/) PIDERIT A.
PA (SUNY/) SUN Y.
XX
PI Macina RA, Piderit A, Sun Y;
XX
DR WPI; 2002-507213/54.
DR N-PSDB; AAL43636.
XX
PT Diagnosing, monitoring, staging, imaging and treating cancers, e.g.
PT gastrointestinal cancers such as stomach, small intestine and colon
PT cancer, associated with the expression of gastrointestinal specific
PT genes Cln114 and Cln115 -
XX
PS Claim 7; Page 17; 23pp; English.
XX
CC The invention comprises a method for diagnosing the presence of
CC gastrointestinal cancers (e.g. cancers of the stomach, small intestine
CC and colon) associated with two gastrointestinal specific genes (GSGs).
CC The two GSGs are human galectin-4 (Cln114) and human carbonic anhydrase I
CC (Cln115). It has been found that Cln114 and Cln115 serve as useful
CC markers in the diagnosis of gastrointestinal cancer. The method of the
CC invention is useful for detecting, diagnosing, monitoring, staging,
CC prognosticating, imaging and treating gastrointestinal cancers associated
CC with the expression of GSGs Cln114 and Cln115. The present amino acid
CC sequence represents the human galectin-4 (Cln114) protein.
XX
SQ Sequence 323 AA;

```
  Query Match             100.0%;  Score 1759;  DB 23;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-158;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323
```

```
RESULT 7
AAU76211
ID   AAU76211 standard; Protein; 323 AA.
XX
AC   AAU76211;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human colon specific gene 9 (CSG9) protein.
XX
KW   Human; colon specific gene; CSG; cytostatic; colon disorder; immunogen;
KW   gastric cancer; ulcerative colitis; granulomatous colitis; colon cancer;
KW   cancer cell.
XX
OS   Homo sapiens.
XX
PN   US6337195-B1.
XX
PD   08-JAN-2002.
XX
PF   31-MAR-1998;   98US-0224110.
XX
PR   06-JUN-1995;   95US-0469667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Yu G,  Rosen C;
XX
DR   WPI; 2002-163239/21.
DR   N-PSDB; ABK15464.
XX
PT   Human colon specific gene polypeptide, useful as diagnostic marker,
PT   vaccine and for screening agonists and antagonists for treating colon
PT   cancer -
XX
PS   Example 1; Fig 9; 49pp; English.
XX
CC   This invention relates to the cDNA and protein sequences of novel
CC   human colon specific genes. These genes which were isolated from a human
CC   colon tissue cDNA library may have cytostatic activity and may be used
CC   as a vaccine against colon disorders such as gastric cancer, ulcerative
CC   colitis and granulomatous colitis. The nucleotide sequences of the
CC   invention may be used to identify corresponding full length gene
CC   sequences and is useful as diagnostic marker for colon cancer and as a
CC   colon cancer vaccine. The protein sequences of the invention are
CC   useful for preparing monoclonal antibodies specific to CSG proteins,
CC   these antibodies can be used to isolate the protein from colon tissue or
CC   for protein quantitation from a sample using enzyme linked immunosorbant
CC   assay (ELISA). It is also useful for targeting cancer cells and for
CC   screening agonist and antagonist for the CSG proteins which are useful
CC   for treating colon cancer. The present sequence represents the human
CC   colon specific gene sequence (CSG9) protein of the invention.
XX
SQ   Sequence   323 AA;
```

```
  Query Match             100.0%;  Score 1759;  DB 23;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-158;
  Matches  323;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323
```

```
RESULT 8
AAO15241
ID   AAO15241 standard; Protein; 329 AA.
XX
AC   AAO15241;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Human galectin-4 (Cln114) protein with a histidine purification tag.
XX
KW   Human; gastrointestinal cancer; stomach cancer; small intestine cancer;
KW   colon cancer; gastrointestinal specific gene; GSG; galectin-4; Cln114;
KW   carbonic anhydrase I; Cln115; gastrointestinal cancer marker;
KW   histidine tag.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 324..329
FT                   /note= "This region represents a 6 histidine purification
FT                   tag"
XX
PN   US2002042088-A1.
XX
PD   11-APR-2002.
XX
PF   09-MAR-2001; 2001US-0802674.
XX
PR   09-MAR-2000; 2000US-188061P.
XX
PA   (MACI/) MACINA R A.
PA   (PIDE/) PIDERIT A.
```

PA (SUNY/) SUN Y.
XX
PI Macina RA, Piderit A, Sun Y;
XX
DR WPI; 2002-507213/54.
XX
PT Diagnosing, monitoring, staging, imaging and treating cancers, e.g.
PT gastrointestinal cancers such as stomach, small intestine and colon
PT cancer, associated with the expression of gastrointestinal specific
PT genes Cln114 and Cln115 -
XX
PS Example 2; Page 15; 23pp; English.
XX
CC The invention comprises a method for diagnosing the presence of
CC gastrointestinal cancers (e.g. cancers of the stomach, small intestine
CC and colon) associated with two gastrointestinal specific genes (GSGs).
CC The two GSGs are human galectin-4 (Cln114) and human carbonic anhydrase I
CC (Cln115). It has been found that Cln114 and Cln115 serve as useful
CC markers in the diagnosis of gastrointestinal cancer. The method of the
CC invention is useful for detecting, diagnosing, monitoring, staging,
CC prognosticating, imaging and treating gastrointestinal cancers associated
CC with the expression of GSGs Cln114 and Cln115. The present amino acid
CC sequence represents the human galectin-4 (Cln114) protein with a
CC C-terminal 6 histidine purification tag.
XX
SQ Sequence 329 AA;

Query Match 100.0%; Score 1759; DB 23; Length 329;
Best Local Similarity 100.0%; Pred. No. 2.7e-158;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
          |||||||||||||||||||||||
Db     301 LSAFQRVDTLEIQGDVTLSYVQI 323
```

RESULT 9
AAG73728
ID AAG73728 standard; Protein; 351 AA.
XX
AC AAG73728;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:4492.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH33159.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 6292-6294; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 351 AA;

Query Match 99.8%; Score 1755; DB 22; Length 351;
Best Local Similarity 99.7%; Pred. No. 7e-158;
Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDV 60
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      29 MAYVPAPGYQPTYNPTLPYYQPIPGGLNVGMSVYIQGVVSEHMKRFFVNFVVGQDPGSDV 88

Qy      61 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 AFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNP 148

Qy     121 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149 FYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQLNSLP 208

Qy     181 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 TMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALHINPR 268

Qy     241 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 MGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLFDFAHR 328

Qy     301 LSAFQRVDTLEIQGDVTLSYVQI 323
          |||||||||||||||||||||||
Db     329 LSAFQRVDTLEIQGDVTLSYVQI 351
```

RESULT 10
AAW61367
ID AAW61367 standard; Protein; 280 AA.

```
XX
AC   AAW61367;
XX
DT   05-OCT-1998  (first entry)
XX
DE   Human galectin-4.
XX
KW   Human; galectin-4; hG4; detection; tumour; breast cancer; diagnosis;
KW   cell proliferative disorder; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO9822139-A1.
XX
PD   28-MAY-1998.
XX
PF   19-NOV-1997;   97WO-US21807.
XX
PR   22-NOV-1996;   96US-0031711.
XX
PA   (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.
XX
PI   Huflejt ME,  Liu F;
XX
DR   WPI; 1998-312177/27.
DR   N-PSDB; AAV28088.
XX
PT   Newly isolated human galectin-4 protein - used to develop products
PT   for the diagnosis, study, prevention and treatment of cancers,
PT   particularly breast cancer
XX
PS   Claim 15; Fig 2; 76pp; English.
XX
CC   The present sequence represents human galectin-4 (hG4). Antibodies
CC   which bind the hG4 protein, and probes or primers derived from the
CC   nucleic acid sequence encoding hG4, can be used for the diagnosis,
CC   study, prevention and treatment of hG4-associated disorders,
CC   particularly cancers (especially breast cancer). They can also be
CC   used for determining the presence of metastases and determining the
CC   susceptibility of a subject with a tumour to develop metastases.
XX
SQ   Sequence   280 AA;

  Query Match             83.9%;  Score 1475.5;  DB 19;  Length 280;
  Best Local Similarity   97.5%;  Pred. No. 1.6e-131;
  Matches  273;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      44 KRFFVNFVVGQDPGSDVAFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFEL 103
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 KRFFVNFVVGQDPGSDVAFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFEL 61

Qy     104 VFIVLAEHYKVVVNGNPFYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMM 163
           ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 VFIVMAEHYKVVVNGNPFYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMM 121

Qy     164 PPYPGPGHCHQQLNSLPTMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAI 223
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 PPYPGPGHCHQQLNSLPTMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAI 181

Qy     224 NFKVGSSGDIALHINPRMGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLD 283
           |||||||||||||||  ||||||||||||||||||| ||||||||||||||||||||||||
Db     182 NFKVGSSGDIALHINPH-GNGTVVRNSLLNGSWGSEEKKTTHNPFGPGQFFDLSIRCGLD 240

Qy     284 RFKVYANGQHLFDFAHRLSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||  ||||||||||||||||||||||||
Db     241 RFKVYANGQHLFDFAHPSRAFQRVDTLEIQGDVTLSYVQI 280


RESULT 11
AAG75625
ID   AAG75625 standard; Protein; 268 AA.
XX
AC   AAG75625;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6389.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH35030.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7869-7871; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   268 AA;

  Query Match             82.9%;  Score 1458;  DB 22;  Length 268;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-130;
  Matches  268;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      56 PGSDVAFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVV 115
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 PGSDVAFHFNPRFDGWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVV 60

Qy     116 VNGNPFYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQ 175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VNGNPFYEYGHRLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGPGHCHQQ 120

Qy     176 LNSLPTMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIAL 235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LNSLPTMEGPPTFNPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIAL 180

Qy     236 HINPRMGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLF 295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 HINPRMGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLDRFKVYANGQHLF 240
```

```
Qy     296 DFAHRLSAFQRVDTLEIQGDVTLSYVQI 323
           |||||||||||||||||||||||||||||
Db     241 DFAHRLSAFQRVDTLEIQGDVTLSYVQI 268
```

RESULT 12
AAY56802
ID AAY56802 standard; Protein; 323 AA.
XX
AC AAY56802;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human eosinophil chemotactic factor (ecalectin).
XX
KW Eosinophil chemotactic factor; galectin; chemotaxis inhibitor;
KW inflammatory disorder; eosinophil; allergic disease; asthma; human;
KW allergic rhinitis; atopic dermatitis; ecalectin; galectin.
XX
OS Homo sapiens.
XX
PN WO9962556-A1.
XX
PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99WO-JP02952.
XX
PR 03-JUN-1998; 98JP-0170698.
XX
PA (EFFE-) EFFECTOR CELL INST.
XX
PI Kanegasaki S, Matsumoto R, Hirashima M;
XX
DR WPI; 2000-086860/07.
DR N-PSDB; AAZ46748.
XX
PT Ecalectin and other galectins as eosinophil chemotaxis promoters for
PT screening potential inhibitors of this effect -
XX
PS Disclosure; Page 35-39; 49pp; Japanese.
XX
CC The invention provides compositions for increasing the chemotacticity of
CC eosinophils that contain as active component a galectin. Chemotaxis
CC inhibitors identified by the screening are used as agents for the
CC treatment and prevention of inflammatory disorders caused by the increase
CC of eosinophil content in affected tissues, such as allergic diseases
CC including bronchial asthma, allergic rhinitis and atopic dermatitis. The
CC present sequence represents a human eosinophil chemotactic factor,
CC ecalectin. Ecalectin belongs to the family of galectin and has
CC eosinophil chemotactic activity.
XX
SQ Sequence 323 AA;

```
  Query Match             32.7%;  Score 576;  DB 21;  Length 323;
  Best Local Similarity   40.1%;  Pred. No. 3.3e-46;
  Matches  131;  Conservative   50;  Mismatches  118;  Indels   28;  Gaps    9;

Qy      11 PTYNPTLPYYQPIPGGLNVGMSVYIQG-VASEHMKRFFVNFVVGQDPGSDVAFHFNPRFD 69
           |  :| :|:   | |||  |: : : | | |    || |||  |   |:|:|||||||:
Db       9 PYLSPAVPFSGTIQGGLQDGLQITVNGTVLSSSGTRFAVNFQTGFS-GNDIAFHFNPRFE 67

Qy      70 GWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLP 129
              || || | | || ||||  |||:||  |:| |:|  :  :||:|||  | :| ||:|
Db      68 DGGYVVCNTRQNGSWGPEERKTHMPFQKGMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP 127

Qy     130 LQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGP-----GHCHQ----QLNSLP 180
             |  : |:| :||  |:|       | | : |  | |      |  |    |: | |
Db     128 FHRVDTISVNGSVQLSYISF--------QPPGVWPANPAPITQTVIHTVQSAPGQMFSTP 179

Qy     181 TMEGPPTFNP----PVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALH 236
            :  ||   |    |:|:   : |||   ::|:: | | |: : | ||  : |   || |
Db     180 AI--PPMMYPHPAYPMPFITTILGGLYPSKSILLSGTVLPSAQRFHIN--LCSGNHIAFH 235
```

```
Qy     237 INPRMGNGTVVRNSLLNGSWGSEEKKITHN-PFGPGQFFDLSIRCGLDRFKVYANGQHLF 295
           :|||     ||||: :: ||||||: :    ||  || | : | |     ||  :|||||
Db     236 LNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFSVWILCEAHCLKVAVDGQHLF 295

Qy     296 DFAHRLSAFQRVDTLEIQGDVTLSYVQ 322
           :: |||     :: ||: ||: |::||
Db     296 EYYHRLRNLPTINRLEVGGDIQLTHVQ 322
```

RESULT 13
ABB77853
ID ABB77853 standard; protein; 323 AA.
XX
AC ABB77853;
XX
DT 27-SEP-2002 (first entry)
XX
DE Amino acid sequence of a human protein.
XX
KW Cancer; galectin 9; antibody; skin cancer; melanoma; breast cancer;
KW ovarian cancer; uterus cancer; tumour; prostate cancer; bladder cancer;
KW kidney cancer; thyroid cancer; throat cancer; tongue cancer;
KW upper jaw cancer; esophageal cancer; stomach cancer; colon cancer;
KW lung cancer; liver cancer; gall-bladder cancer; pancreatic cancer;
KW leukemia; liposarcoma; glioma.
XX
OS Homo sapiens.
XX
PN WO200237114-A1.
XX
PD 10-MAY-2002.
XX
PF 31-OCT-2001; 2001WO-JP09561.
XX
PR 01-NOV-2000; 2000JP-0335077.
XX
PA (GALP-) GALPHARMA CO LTD.
XX
PI Hirashima M, Yamauchi A, Kageshita T, Nakamura T, Nishi N;
XX
DR WPI; 2002-519265/55.
XX
PT Metastasis mechanism-based agents (anti-galectin 9 antibody) for
PT detecting ability of cancer to metastasize in cells, uses galectin 9 as
PT marker to detect cancer metastasis for diagnosis, cancer prevention and
PT treatment -
XX
PS Example 5; Page 61-63; 68pp; Japanese.
XX
CC The specification describes an agent for detecting the ability of
CC cancers to metastasize. This agents an comprises anti-galectin 9
CC antibody as the active ingredient. The agents and method are for or
CC detecting ability of cancer to metastasize in cancer cells, which are
CC useful in the diagnosis, prevention and treatment of cancer. The cancer
CC includes epithelial malignant tumours (tumorising or non-tumorising)
CC in organs, tissues or blood. Such cancer can be 1 of the not less than 31
CC specifically claimed, e.g. skin cancer (including melanoma), breast
CC cancer, ovarian cancer, uterus cancer, malignant tumour of the testis,
CC prostate cancer, bladder cancer, kidney cancer, thyroid cancer, cancer
CC of throat and larynx, tongue cancer, upper jaw cancer, esophageal
CC cancer, stomach cancer, colon and rectum cancer, lung and bronchus
CC cancer, liver cancer (including hepatocarcinoma and intrahepatic
CC biliary cancer), extrahepatic biliary duct and gall-bladder cancer,
CC pancreatic cancer, leukemia, malignant lymphoma, liposarcoma and
CC glioma. The cancer cells are particularly breast cancer cells or
CC melanoma cells. The present sequence represents a human protein,
CC which is used in the course of the invention.
XX
SQ Sequence 323 AA;

```
  Query Match             32.7%;  Score 576;  DB 23;  Length 323;
  Best Local Similarity   40.1%;  Pred. No. 3.3e-46;
  Matches  131;  Conservative   50;  Mismatches  118;  Indels   28;  Gaps    9;
```

```
Qy      11 PTYNPTLPYYQPIPGGLNVGMSVYIQG-VASEHMKRFFVNFVVGQDPGSDVAFHFNPRFD 69
           |  :| :|:   | |||  |: : : | | |    || |||  |   |:|:||||||||:
Db       9 PYLSPAVPFSGTIQGGLQDGLQITVNGTVLSSSGTRFAVNFQTGFS-GNDIAFHFNPRFE 67

Qy      70 GWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLP 129
             || || | | || ||||  |||:||  |:| |:| :  :||:|||  | :| ||:|
Db      68 DGGYVVCNTRQNGSWGPEERKTHMPFQKGMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP 127

Qy     130 LQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGP-----GHCHQ----QLNSLP 180
           |  : |:| :||  |:|       | | : |  | |      |  |    |: | |
Db     128 FHRVDTISVNGSVQLSYISF--------QPPGVWPANPAPITQTVIHTVQSAPGQMFSTP 179

Qy     181 TMEGPPTFNP----PVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALH 236
           :  ||   |    |:|:   : |||   ::|:: | | |: : | ||  : |   || |
Db     180 AI--PPMMYPHPAYPMPFITTILGGLYPSKSILLSGTVLPSAQRFHIN--LCSGNHIAFH 235

Qy     237 INPRMGNGTVVRNSLLNGSWGSEEKKITHN-PFGPGQFFDLSIRCGLDRFKVYANGQHLF 295
           :|||    ||||: :: ||||||: :    ||  || | : | |     ||  :|||||
Db     236 LNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFSVWILCEAHCLKVAVDGQHLF 295

Qy     296 DFAHRLSAFQRVDTLEIQGDVTLSYVQ 322
           :: |||     :: ||: ||: |::||
Db     296 EYYHRLRNLPTINRLEVGGDIQLTHVQ 322
```

RESULT 14
ABP61494
ID ABP61494 standard; Protein; 323 AA.
XX
AC ABP61494;
XX
DT 30-SEP-2002 (first entry)
XX
DE Human NF-kB activating protein SEQ ID NO 141.
XX
KW Human; NF-kB; nuclear factor kappa B; mouse; antiinflammatory;
KW immunomodulator; cytostatic; antiinfective; osteopathic; nootropic;
KW neuroprotective; anti-HIV; autoimmune disease; cancer; infection;
KW bone disease; AIDS; neurodegenerative disease; ischaemic disorder.
XX
OS Homo sapiens.
XX
PN WO200253737-A1.
XX
PD 11-JUL-2002.
XX
PF 25-DEC-2001; 2001WO-JP11389.
XX
PR 28-DEC-2000; 2000JP-0402288.
PR 26-MAR-2001; 2001JP-0088912.
PR 24-AUG-2001; 2001JP-0254018.
XX
PA (ASAH) ASAHI KASEI KOGYO KK.
XX
PI Matsuda A, Honda G, Muramatsu S, Nagano Y;
XX
DR WPI; 2002-583617/62.
DR N-PSDB; ABQ91982.
XX
PT NF-approximatelykB activating gene and expressed protein, applicable in
PT diagnosis and screening inhibitors or promoters to control excessive
PT activation or inhibition for treating e.g. inflammations, autoimmune
PT diseases and cancer -
XX
PS Claim 1; Page 613-615; 841pp; Japanese.
XX
CC The invention relates to a purified protein (I), comprising one of 90
CC fully defined sequences (ABP61424-ABP61513) or a protein based on any of
CC the sequences but with some amino acids deleted, substituted or added and
CC with a NF-kB (nuclear factor kappa B) activating effect. The protein and
CC encoding gene (ABQ91912-ABQ92001) are useful in diagnosis and screening
CC inhibitors or promoters to control excessive activation or inhibition
CC and for treating e.g. inflammations, autoimmune diseases, cancers,
CC infections, bone diseases, AIDS, neurodegenerative diseases or ischaemic
CC disorders.
XX
SQ Sequence 323 AA;

Query Match 32.7%; Score 576; DB 23; Length 323;
Best Local Similarity 40.1%; Pred. No. 3.3e-46;
Matches 131; Conservative 50; Mismatches 118; Indels 28; Gaps 9;

```
Qy      11 PTYNPTLPYYQPIPGGLNVGMSVYIQG-VASEHMKRFFVNFVVGQDPGSDVAFHFNPRFD 69
           |  :| :|:   | |||  |: : : | | |    || |||  |   |:|:||||||||:
Db       9 PYLSPAVPFSGTIQGGLQDGLQITVNGTVLSSSGTRFAVNFQTGFS-GNDIAFHFNPRFE 67

Qy      70 GWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGHRLP 129
             || || | | || ||||  |||:||  |:| |:| :  :||:|||  | :| ||:|
Db      68 DGGYVVCNTRQNGSWGPEERKTHMPFQKGMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP 127

Qy     130 LQMVTHLQVDGDLQLQSINFIGGQPLRPQGPPMMPPYPGP-----GHCHQ----QLNSLP 180
           |  : |:| :||  |:|       | | : |  | |      |  |    |: | |
Db     128 FHRVDTISVNGSVQLSYISF--------QPPGVWPANPAPITQTVIHTVQSAPGQMFSTP 179

Qy     181 TMEGPPTFNP----PVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAINFKVGSSGDIALH 236
           :  ||   |    |:|:   : |||   ::|:: | | |: : | ||  : |   || |
Db     180 AI--PPMMYPHPAYPMPFITTILGGLYPSKSILLSGTVLPSAQRFHIN--LCSGNHIAFH 235

Qy     237 INPRMGNGTVVRNSLLNGSWGSEEKKITHN-PFGPGQFFDLSIRCGLDRFKVYANGQHLF 295
           :|||    ||||: :: ||||||: :    ||  || | : | |     ||  :|||||
Db     236 LNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFSVWILCEAHCLKVAVDGQHLF 295

Qy     296 DFAHRLSAFQRVDTLEIQGDVTLSYVQ 322
           :: |||     :: ||: ||: |::||
Db     296 EYYHRLRNLPTINRLEVGGDIQLTHVQ 322
```

RESULT 15
AAB97171
ID AAB97171 standard; Protein; 341 AA.
XX
AC AAB97171;
XX
DT 13-AUG-2001 (first entry)
XX
DE Rainbow trout galectin.
XX
KW Rainbow trout; galectin; cytostatic; antiinflammatory; immunity;
KW fish disease; cell adhesion inhibition; inflammation; cancer;
KW T cell movement; apoptosis.
XX
OS Oncorhynchus mykiss.
XX
PN JP2001069976-A.
XX
PD 21-MAR-2001.
XX
PF 01-SEP-1999; 99JP-0247204.
XX
PR 01-SEP-1999; 99JP-0247204.
XX
PA (NORQ) NORINSUISANSHO YOSHOKU KENKYU.
XX
DR WPI; 2001-321173/34.
DR N-PSDB; AAH24641.
XX
PT Novel recombinant Rainbow trout galectin protein and gene encoding the
PT protein, useful for studying fish immunity mechanism, and diagnosis of
PT fish diseases -
XX
PS Claim 1; Page 10-11; 17pp; Japanese.
XX
CC The invention relates to a recombinant protein having a 341 amino acid
CC sequence fully defined in the specification, or its mutant in which at
CC least one amino acid is deleted, substituted or added, but which retains

```
CC   the galectin activity. The protein is useful for studying mechanisms of
CC   fish immunity and for diagnosing fish diseases. The galectin proteins are
CC   involved in inhibition of cell adhesion, inflammation, metastasis of
CC   tumour cells, T cell movement, and apoptosis. The present sequence
CC   is the protein of the invention.
XX
SQ   Sequence   341 AA;

  Query Match             32.7%;  Score 575.5;  DB 22;  Length 341;
  Best Local Similarity   39.7%;  Pred. No. 4e-46;
  Matches  135;  Conservative   38;  Mismatches  136;  Indels   31;  Gaps    8;

Qy      10 QPTYNPTLPYYQPIPGGLNVGMSVYIQGVASEHMKRFFVNFVVGQDPGSDVAFHFNPRFD 69
           :| |||:||:   | |||  | :: | |       || ||   |   |:|:| |||||:|
Db       7 EPFYNPSLPFTGSIQGGLQEGKTITITGRVHHGADRFHVNLQCGSRAGADIALHFNPRYD 66

Qy      70 ---GWDKVVFNTLQGGKWGSEERKRSMPFKKGAAFELVFIVLAEHYKVVVNGNPFYEYGH 126
              |:  || ||||  |||:||||:  ||::|::| |   |  : ::: |||| |  : |
Db      67 SHPGY--VVTNTLQQSKWGTEERKQHSPFQRGSSFSLEITVQRDFFQLKVNGNHFMTFKH 124

Qy     127 RLPLQMVTHLQVDGDLQLQSINFIGGQPLRPQGP-----PMMPPYPG----------PGH 171
           |:|   |  :  || ::| || |    |  | |     |  | |||          ||
Db     125 RIPFYSVDTISADGKVELTSIVFQNPAPTIPAQPGFPAQPGFPSYPGFPAQPGFPSYPGF 184

Qy     172 CHQQ-LNSLPTMEGPPTF-------NPPVPYFGRLQGGLTARRTIIIKGYVPPTGKSFAI 223
             |   | |   | | |       | |||   : |||   ||| |:| | |    | |
Db     185 PAQPGFPSCPGFPGQPGFPYPGFPAQPAVPYKNMINGGLYPGRTINIQGVVNPNANRFHI 244

Qy     224 NFKVGSSGDIALHINPRMGNGTVVRNSLLNGSWGSEEKKITHNPFGPGQFFDLSIRCGLD 283
           |    |   |||| |||     |||| |   || ||:     ||  || | ||| |
Db     245 NLLFNSG--IALHFNPRFDETLVVRNSKLRDQWGKEERS-GGMPFHRGQAFTLSITCDAQ 301

Qy     284 RFKVYANGQHLFDFAHRLSAFQRVDTLEIQGDVTLSYVQI 323
           :|:  ||     : || :  |:|: ||: ||::|: | :
Db     302 CYKIVVNGNQTSTYKHRHTLLQQVNILEVDGDLSLTSVMV 341


Search completed: January  7, 2003, 16:17:46
Job time : 39 secs
```